"戴玉数据圈"沙龙：央视大数据新闻的开创和探索

文 / 郭俊义　戴　玉　薛　葵　高　赫

编者按："戴玉数据圈"是资深数据新闻人戴玉创办的数据新闻群，聚集了中国最顶尖的数据新闻媒体主编和团队负责人，汇聚各大智库、数据源公司的对接人，以及各大高校教授数据新闻的老师。中国的这些数据新闻人，正在开创什么样的报道新形式？群内定期举办最前沿的沙龙，揭秘中国数据新闻新动向。

chinaXiv:202310.02886v1

主持人：戴玉

《南风窗》原编辑、人民日报原中央厨房数据新闻项目负责人。率团队获得中国首届数据新闻大赛二等奖，入围2016英国信息之美设计大赛长名单。



戴　玉

嘉宾：郭俊义

央视新闻联播编辑部副制片人、中国大数据新闻创始人。



郭俊义

郭俊义：

各位数据圈的朋友，大家好！今天我主要想介绍几个央视大数据报道的开创之作和经典案例，然后再聊聊我对大数据新闻未来的一些想法。

1.《据说春运》系列

相信大家比较了解的是我们最早跟百度合作的《据说春运》系列。很多人都记得央视播放的百度春运迁徙图，但并不知道这个新闻是怎么做出来的，今天可以给大家揭秘一下。

当时是2014年1月，临近春节，百度开发了一个"百度春运迁徙图"的产品。在3年前大家可能会觉得这个迁徙图很酷炫，但它是一款可视化产品，还不是新闻作品。我特别强调一下，一个可视化产品是不具有新闻性的，也就不会吸引大家来看。

但究竟怎么从大数据里面挖掘出新闻？当时也是非常费脑筋。百度并没有给我们提供新闻点，所以全靠我们自己挖掘。当时我从百度春运迁徙图里面仔细看了很多城市的人员流入流出、热门线路等，花了很长时间发现了一个新闻点。

大家知道，中国春运以往的特点在年前主要是从特大城市向中小城市迁徙，东南沿海向中西部城市迁徙，但这种常规规律就不足以构成一个新闻点。当时我发现，北京迁往成都的迁徙量非常大，但从成都迁往北京的流出量也非常大，连续几天都是排行前三名。我们就从这个数据发现了一个春运新变化——出现了逆向迁徙，即从中西部城市向特大城市迁徙。当然，这仅仅是从百度给我们提供的手机定位大数据来分析的。

因为当时第一次做大数据新闻，我们从一家商业公司给我们提供的迁徙大数据得出了一个新闻点，心里并不踏实。

ChinaXiv合作期刊

chinaXiv:202310.02886v1

作为国家媒体央视来说，这么轻率地发布是不是有权威性？所以我们当时还利用了传统的调查方式，请我们四川记者站的记者采访了成都铁路局的工作人员。发现他们得到的数据也支撑了我们的新闻点，即从实际售票的情况来讲，成都的迁出量确实非常大。

即便是基于成都铁路局提供的售票数据，我们当时觉得还是不踏实。因为电视的话，需要有一个直观的画面来证实我们的新闻点。我们就请成都火车站的工作人员给我们拍了一些成都火车站候车大厅的现场实况。这个现场实况当时也非常符合我们的新闻点。因为在画面中，比较明显的是有很多老人要出行。当时还有 6 天就要除夕了，老人为什么要出行？他们不可能出去打工也不可能出去旅游，更多的可能是逆向迁徙到子女所在的城市过年。

所以我们第一期节目的内容就是通过交叉印证的方法确定了新闻点，支撑了我们的内容。接下来就是如何把它变成比较酷炫的可视化产品搬上屏幕。当时想到很多办法，比如翻拍电脑、用笔记本电脑直接接入演播室大屏等，但是效果都不好。后来想了一个办法——现在看来也是比较简单的办法，就是通过录屏的方式。我们直接用录屏软件把百度迁徙图 flash 格式的动态图转换为一个视频格式，就可以直接在电视屏幕上呈现了。

我们第一期节目成功地推出后，影响是非常大的。因为很多观众也好，学界业界的朋友也好，从来没看过这种形式，所以反响比较大，我们也比较受鼓舞。后来我们接连做了很多个以《据说》为名称的系列报道，有据说春运、据说春节、据说两会、据说世界杯、据说 APEC 等。

2. 大数据人群画像

再跟大家分享一个我自己觉得做得比较好的一期节目。这期节目就是我们跟 360 公司合作的一期“两会”报道。当时比较重要的一个热点就是网络诈骗，我们与 360 公司的首席安全员裴博士商量怎么用 360 的安全大数据来做相关报道。

裴博士问我说，他们有很多安全大数据，是不是可以从中选取。但是，这些技术性很强的大数据并不能成为新闻，就是说我们媒体自己还要在选题上想办法。后来我就想到一个角度，请裴博士用 360 安全大数据做一个人群画像，给我描述一下什么样的人上网最容易上当受骗。360 公司就花了两三天时间做了这个画像。

画像的结果是，生活在广东的，16 岁到 25 岁，即“90 后”男孩，在上午 10 点到 12 点时，从事网络兼职、网上购物和网络游戏等几项活动，就很容易被骗。画像出来之后我们觉得很有意思，但它毕竟只是从大数据得来的人群特征，现实中是不是真的是这样呢？

当时也是很凑巧，在做出人群画像的前一天，中山大学的官方微博就发了一条微博，说近期本校有多位男生，因为上网找工作、找兼职等上当受骗。这个案例我们一看到，就发现与大数据挖掘出来的人群画像非常吻合，所以说大数据真的很神奇。

3. 上海外滩踩踏事件

我还想跟大家分享一个案例，上海外滩的踩踏事件。大家都知道，2014 年 12 月 31 号，上海外滩发生了踩踏惨剧，当时踩踏死亡 36 人。第二天我听说了这个惨剧之后想到，大数据有预警功能，我们能不能使用大数据来对相关的大型活动人群聚集做一个预警。

人流密度，我们现在叫热力图，其实它也就是来源于基于手机定位的 LBS 大数据。当时我找到腾讯地图，把我的这个需求跟他们交流了一下，他们配合我做了一个踩踏事故当天、前一天和后一天上海外滩的人流变化热力图，还有一张人流变化趋势图。

这两个图都能够非常明显地看到人流随时间而发生的变化。如果说我们上海外滩所在的黄浦区的相关部门能够早做大数据预警的话，这个惨剧就有可能被避免。

挖掘这种人流聚集的防踩踏的大数据，我觉得很有现实意义，所以后来我就跟中国移动的相关领导聊了这么一个想法。中国移动后来请北京移动公司专门做了一个北京庙会的防踩踏的软件，2015 年春节应用在地坛庙会，作为防踩踏预警，效果很好。

4. 数据微述评

接下来我想讲一下我们最近的一个新的做法。新闻联播在 2016 年底做了一个年度报道叫《点点改革年度账》。这个系列报道做了八集。在这里我们有个新的探索，可以跟大家分享一下。

《点点改革年度账》这个系列报道，前面一个“点”是盘点，后面一个“点”是点评。盘点就是用数据来盘点改革的措施、成效等；点评就是我们每一个数据都附上一句话点评。我们后来把它总结为“数据微述评”，这个形式在新闻报道中比较少见，算是有一点新的突破。

做了 3 年数据新闻后，我有些体会和想法。大数据新闻有门槛，需要大数据来源的同时，也需要数据工程师、数据分析师等专业性比较强的专业人员，需要一个团队去操作。未来希望与相关大数据公司合作，组成大数据新闻团队，挖掘出有新闻价值和实际价值的数据新闻产品。

主持人戴玉：

非常感谢郭老师给我们带来的鲜活案例。央视将大数据分析、现场采访调查、数据可视化和画面语言结合到一起，做出来的产品综合性很强，影响力也很广。那么我想问一下，您觉得电视媒体做数据新闻跟其他形式、媒体相比有什么优势和劣势呢？

chinaXiv:202310.02886v1

郭俊义：

我觉得电视媒体的优势就是视觉效果吧，电视可以运用很丰富的多种动态可视化效果去呈现包括数据在内的内容，做出来的东西很好看，也很直观。但劣势就是深度报道的推理在电视节目中不容易展开，而且一期数据新闻的制作成本也比较高。

戴玉：

您刚才提到了成本的问题。有的圈友也很好奇，央视和大数据公司合作生产数据新闻，是否需要向数据公司付费？

郭俊义：

我们的成本主要还是花在后期制作上面了，跟大数据公司的合作不涉及经费。新闻媒体做数据新闻的第一考量肯定是新闻价值，我们要新闻价值，大数据公司要传播价值，双方都有合作需求。当然，央视的传播平台大，也是一个吸引公司合作的重要因素。

戴玉：

也有大数据公司向我抱怨过，虽然跟媒体合作确实能挖掘出很多意料之外的结果，但对于大数据公司来说，调用数据和人力好不容易得出来的一些数据结论，未被媒体采用的话就浪费掉了，他们也会考虑数据分析成本和媒体传播效果之间的性价比。

郭俊义：

这可以理解，但作为媒体来说，新闻价值是第一选择，媒体的版面也是有限的呀！大数据新闻的投入是比较大，百度跟我们合作，最多时安排了20多人的团队来配合，但选题的社会影响力也相当可观。媒体报道之外的那些数据结论，大数据公司也可以放到自己的公众号上去发布。当然，话说回来，这可能需要媒体和数据公司更好地沟通。如果记者提出的选题纬度需求更精准，就可以避免数据结论的浪费。

戴玉：

的确，好的数据记者需要能够驾驭大数据为新闻所用，有时候解读上的偏差也会造成“误报”。刚才群里也有社科院的老师提出来，一些学者并不认可数据的科学意义，仍然认为只有传统的社会田野调查才更真实一些，不知道您怎么看待大数据新闻的科学性？

郭俊义：

我曾跟一个法学教授探讨过运用大数据研究法学，他们没听说过，后来真有运用大数据研究法学的相关论文出现。我不是社会学家，但我知道传统的社会调查都是抽样调查，而大数据是全样本，我觉得全样本更有价值。但我也不是统计学家，也不敢否认抽样调查的价值，但是既然有了大数据及挖掘分析能力，为什么不能尝试分析一下海量的大数据呢？我觉得对大数据来源的选择，一是要权威，在各自领域领先，另外媒体可以利用传统调查、交叉印证等手法规避数据来源的片面性。

戴玉：

数据科学家的数据挖掘，再加上记者的实地调查采访，或许这种综合研究和多面核实正是未来数据界和新闻界共同的趋势吧。媒

